System Aware Cyber Security

Application of Dynamic System Models and State Estimation Technology to the Cyber Security of Physical Systems

Barry M. Horowitz, Kate Pierce

University of Virginia

April, 2012

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract H98230-08-D-0171. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology

Objectives for System Aware Cyber Security Research

- Increase cyber security by developing new system engineering-based technology that provides a Point Defense option for cyber security
    - Inside the system being protected, for the most critical functions
    - Complements current defense approaches of network and perimeter cyber security
- Directly address supply chain and insider threats that perimeter security does not protect against
    - Including physical systems as well as information systems
- Provide technology design patterns that are reusable and address the assurance of data integrity and rapid forensics, as well as denial of service
- Develop a systems engineering scoring framework for evaluating cyber security architectures and what they protect, to arrive at the most cost-effective integrated solution

Publications

Jennifer L. Bayuk and Barry M. Horowitz, An Architectural Systems Engineering Methodology for Addressing Cyber Security, Systems Engineering 14 (2011), 294-304.

- Rick A. Jones and Barry M. Horowitz, System-Aware Cyber Security, ITNG, 2011 Eighth IEEE International Conference on Information Technology: New Generations, April, 2011, pp. 914-917. (Best Student Paper Award)
- Rick A. Jones and Barry M. Horowitz, System-Aware Security for Nuclear Power Systems, 2011 IEEE International Conference on Technologies for Homeland Security, November, 2011. (Featured Conference Paper)
- Rick A. Jones and Barry M. Horowitz, A System-Aware Cyber Security Architecture, Systems Engineering, Volume 15, No. 2, February, 2012

System-Aware Cyber Security Architecture

- System-Aware Cyber Security Architectures combine design techniques from 3 communities
  - Cyber Security
  - Fault-Tolerant Systems
  - Automatic Control Systems
- The point defense solution designers need to come from the communities related to system design, providing a new orientation to complement the established approaches of the information assurance community
- New point defense solutions will have independent failure modes from traditional solutions, thereby minimizing probabilities of successful attack via greater defense in depth

A Set of Techniques Utilized in System-Aware Security

Cyber Security

*Data Provenance

*Moving Target
(Virtual Control for Hopping)

*Forensics

Fault-Tolerance

*Diverse Redundancy
(DoS, Automated Restoral)

*Redundant Component Voting
(Data Integrity, Restoral)

Automatic Control

*Physical Control for
Configuration Hopping
(Moving Target, Restoral)

*State Estimation
(Data Integrity)

*System Identification
(Tactical Forensics, Restoral)

A Set of Techniques Utilized in System-Aware Security

Cyber Security

*Data Provenance

*Moving Target

(Virtual Control for Hopping)

*Forensics

Fault-Tolerance

*Diverse Redundancy

(DoS, Automated Restoral)

*Redundant Component Voting

(Data Integrity, Restoral)

Automatic Control

*Physical Control for

Configuration Hopping

(Moving Target, Restoral)

*State Estimation

(Data Integrity)

*System Identification

(Tactical Forensics, Restoral)

This combination of solutions requires adversaries to:

- Understand the details of how the targeted systems actually work

A Set of Techniques Utilized in System-Aware Security

Cyber Security

*Data Provenance

*Moving Target
(Virtual Control for Hopping)

*Forensics

Fault-Tolerance

*Diverse Redundancy
(DoS, Automated Restoral)

*Redundant Component Voting
(Data Integrity, Restoral)

Automatic Control

*Physical Control for
Configuration Hopping
(Moving Target, Restoral)

*State Estimation
(Data Integrity)

*System Identification
(Tactical Forensics, Restoral)

This combination of solutions requires adversaries to:

- Understand the details of how the targeted systems actually work
- Develop synchronized, distributed exploits consistent with how the attacked system actually works

A Set of Techniques Utilized in System-Aware Security

Cyber Security

*Data Provenance

*Moving Target
(Virtual Control for Hopping)

*Forensics

Fault-Tolerance

*Diverse Redundancy
(DoS, Automated Restoral)

*Redundant Component Voting
(Data Integrity, Restoral)

Automatic Control

*Physical Control for
Configuration Hopping
(Moving Target, Restoral)

*State Estimation
(Data Integrity)

*System Identification
(Tactical Forensics, Restoral)

If implemented properly, this combination of solutions requires adversaries to:

- Understand the details of how the targeted systems actually work
- Develop synchronized, distributed exploits consistent with how the attacked system actually works
- Corrupt multiple supply chains

Example Design Patterns Under Development

- **Diverse Redundancy** for post-attack restoration
- **Diverse Redundancy + Verifiable Voting** for trans-attack defense
- **Physical Configuration Hopping** for moving target defense
- **Virtual Configuration Hopping** for moving target defense
- **Physical Confirmations of Digital Data**
- **Data Consistency Checking**

ATTACK 1: OPERATOR DISPLAY ATTACK

ATTACK 2: CONTROL SYSTEM &
OPERATOR DISPLAY ATTACK

ATTACK 3: SENSOR SYSTEM ATTACK

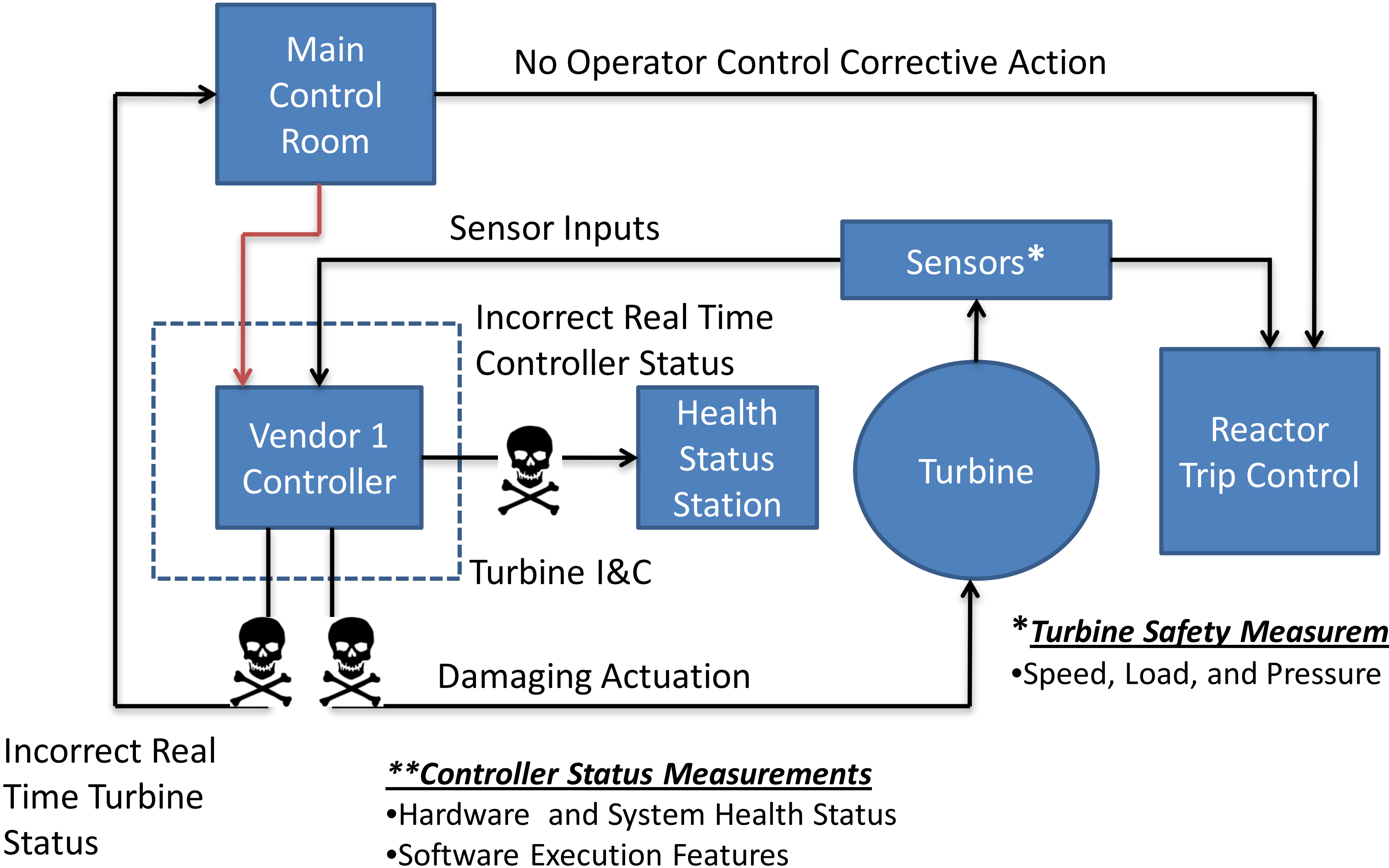
ATTACKS 1 & 2

OPERATOR DISPLAY ATTACK/ COORDINATED CONTROL SYSTEM & OPERATOR DISPLAY ATTACK

The Problem Being Addressed

- Highly automated physical system
- Operator monitoring function, including criteria for human over-ride of the automation
- Critical system states for both operator observation and feedback control – consider as *least trusted from cyber security viewpoint*
- Other measured system states – consider as *more trusted from cyber security viewpoint*
- CYBER ATTACK: Create a problematic outcome by disrupting human display data and/or critical feedback control data.

Cyber Attack: Damaging Turbine and Hiding its Effects



Main Control Room

No Operator Control Corrective Action

Sensor Inputs

Sensors*

Incorrect Real Time Controller Status

Vendor 1 Controller

Health Status Station

Turbine

Reactor Trip Control

Turbine I&C

Damaging Actuation

Incorrect Real Time Turbine Status

*<u>***Turbine Safety Measurements***</u>

•Speed, Load, and Pressure

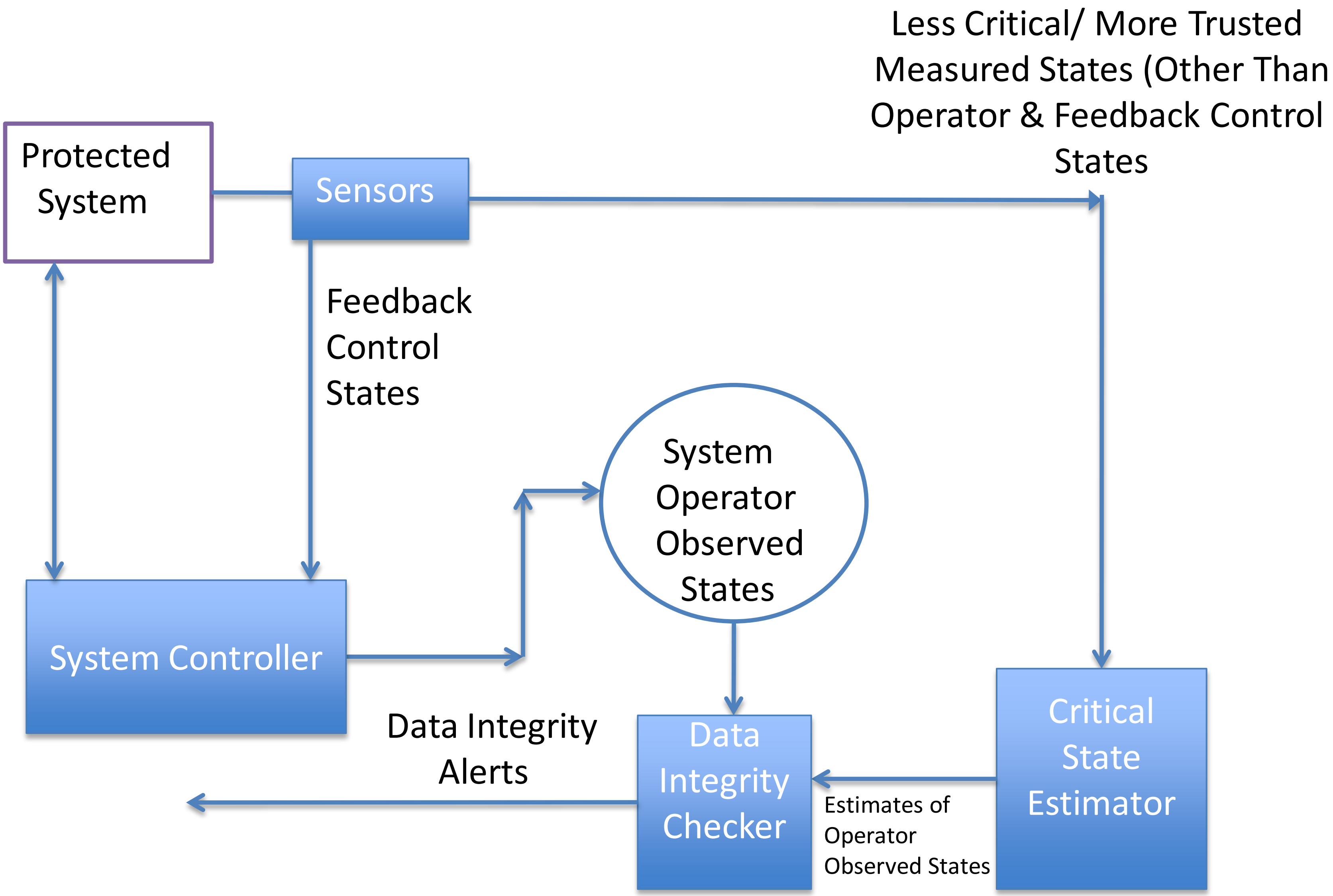
<u>*****Controller Status Measurements***</u>

•Hardware and System Health Status

•Software Execution Features

•I/O Status

Simplified Block Diagram for Inference-Based Data Integrity Detection System


Less Critical/ More Trusted Measured States (Other Than Operator & Feedback Control States
Protected System
Sensors
Feedback Control States
System Operator Observed States
System Controller
Data Integrity Alerts
Data Integrity Checker
Estimates of Operator Observed States
Critical State Estimator

EXAMPLE

Regulating a Linear Physical System (1)

- Linear physical system represented by difference equation
- x (k+1)=Ax (k)+Bu (k)+ω (k) where
- x (k) is an n vector representing the system state during discrete time interval k
- A is the n x n system state transition matrix
- B is the n x g system control matrix
- u (k) is the g vector control signal
- ω (k) is system input noise

Regulating a Linear Physical System (2)

- System measurements are represented by:
- y (k) = C x (k) + v (k)
- Where y(k) is a m vector of measurements at time interval k
- C is a mxn measurement matrix
- v (k) is an m vector representing measurement noise

A Simulation Model for Regulating the States of the System

- To facilitate evaluating the data consistency cyber security design pattern:
  - Simulate a linear system controller to sustain the states of a system at designated levels
  - Optimal Regulator Solution (LQG) utilized for simulation
    - White Gaussian noise
    - Separation Theorem
    - Kalman Filter for state estimation
    - Ricatti Equation-based controller for feedback control
  - Controller feed back law based upon variances of input noise, measurement noise and the A,B and C matrices of the system dynamics model

Example State Equations and Noise Assumptions

```
A = [ 1,   1.  -.02,  -.01
       .01,  1,  -.01,   0
       .2,  .01,    1,     1
     -.01, .02, -.01,  1 ];
B = [ 0 ,  1 , 0 , 0 ];
```

Operator Observed (less trusted):

C = [1, 0, 0, 0];

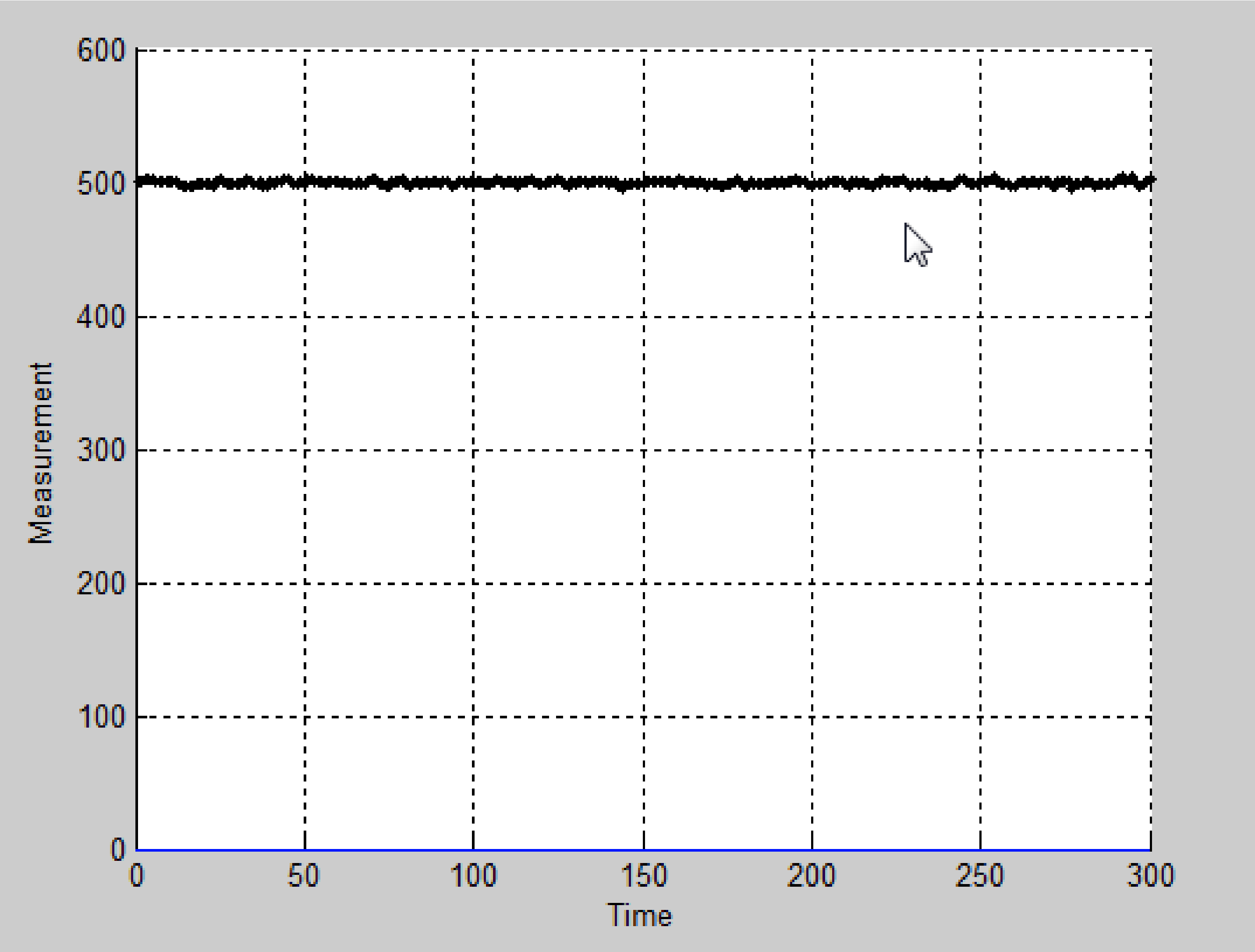
Related States (unobserved by operator, more trusted):

C2 = [0 1 0 0; 0 0 1 0; 0 0 0 1]

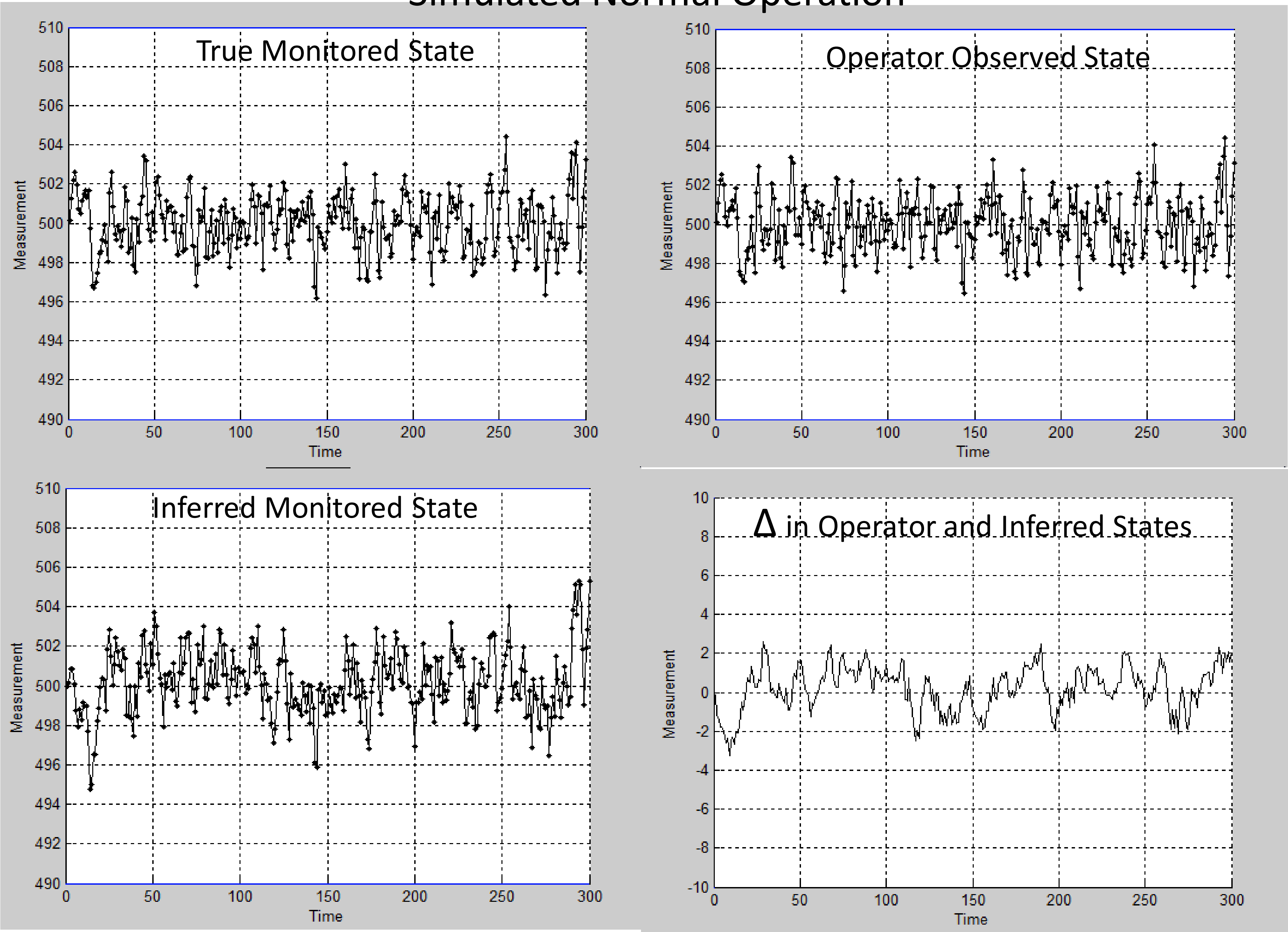
K1 = 0.25;   process noise variances for each of the states

K2 = 0.25;   sensor noise variances for each of the measurements

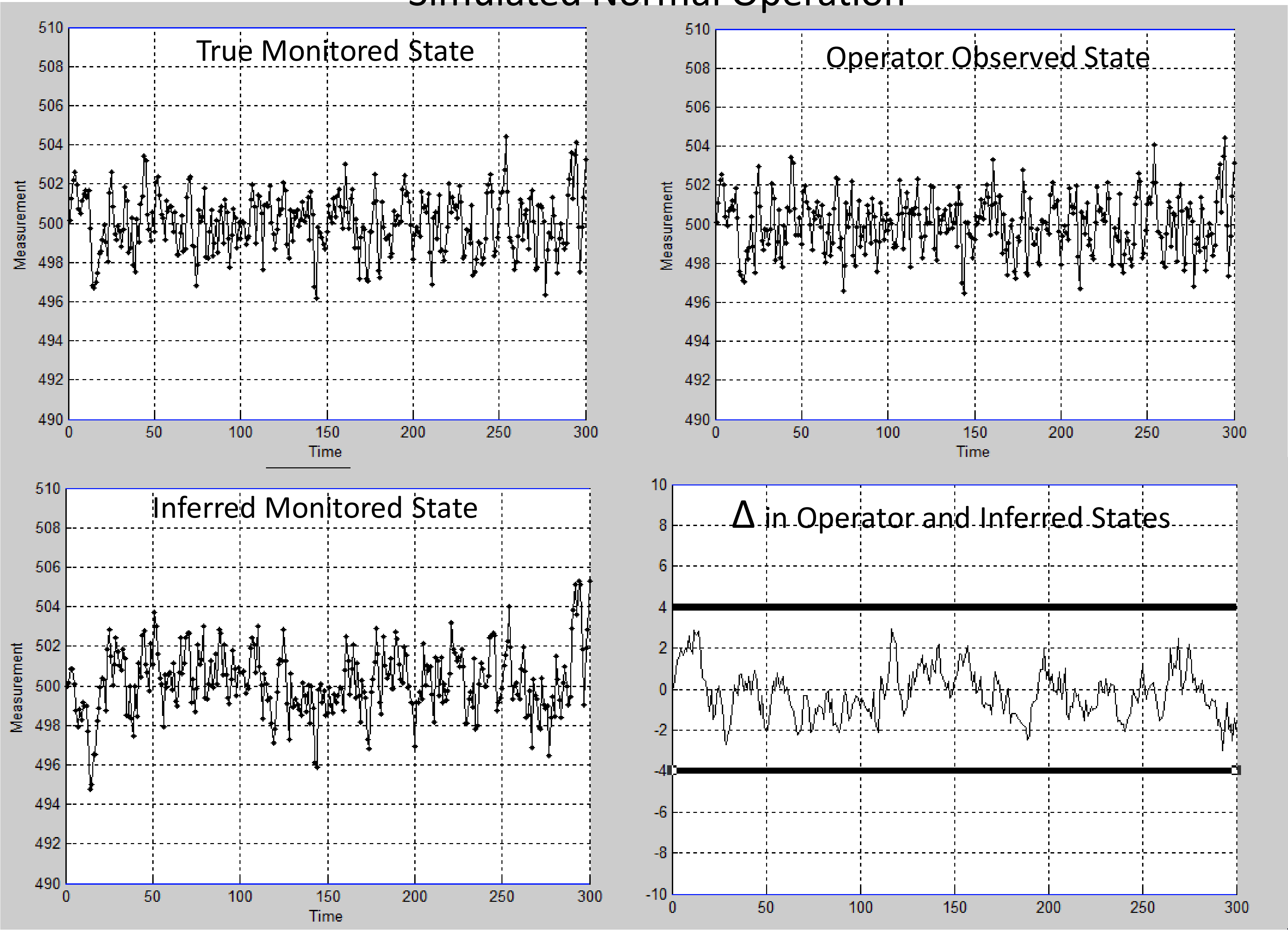
Simulated System Operation for Regulation of a State Component at 500



Simulated Normal Operation



Simulated Normal Operation



True Monitored State

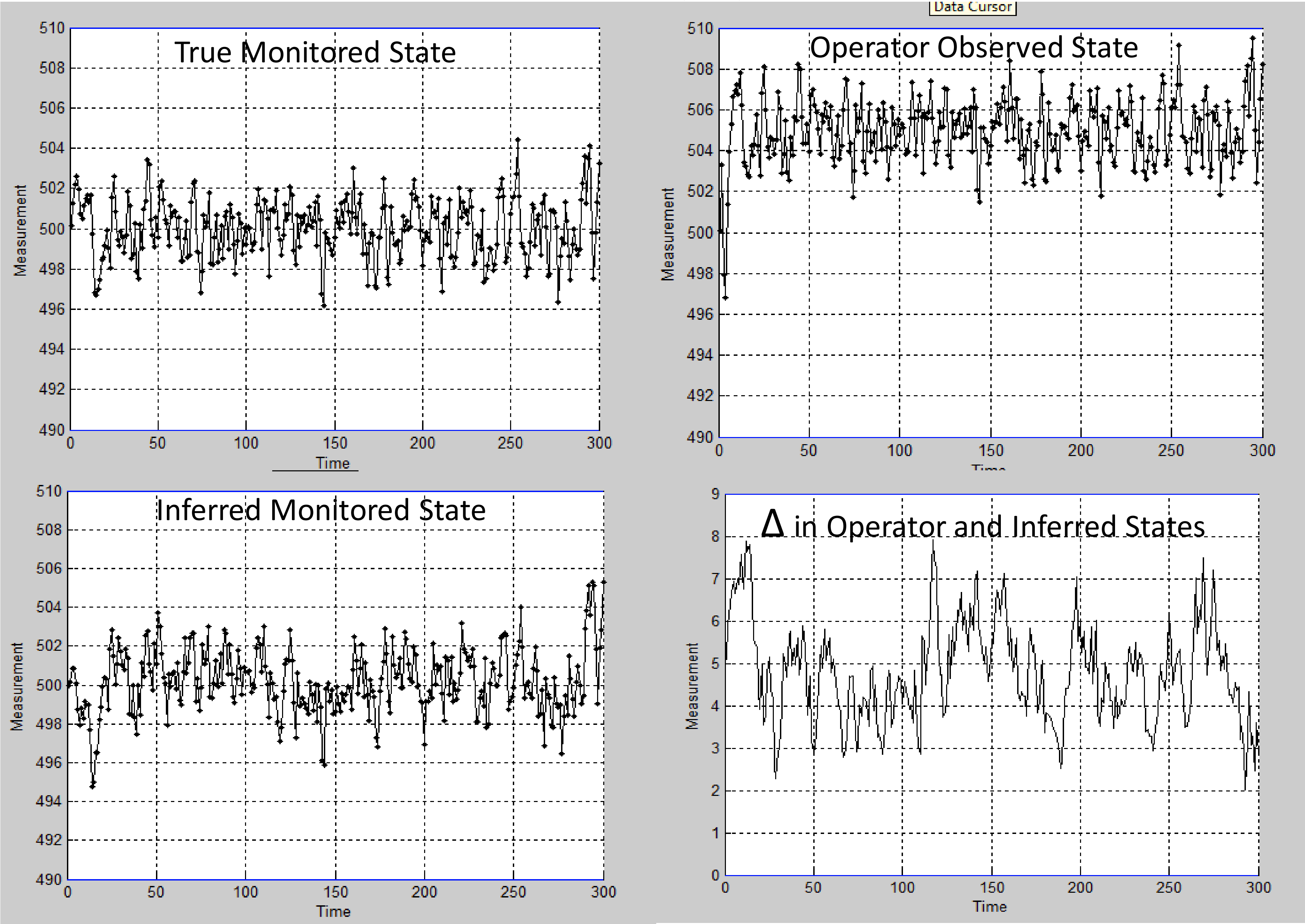
Operator Observed State

Inferred Monitored State

Δ in Operator and Inferred States

REPLAY ATTACK TO CAUSE ERRONEOUS OPERATOR ACTION

Simulated Replay Attack



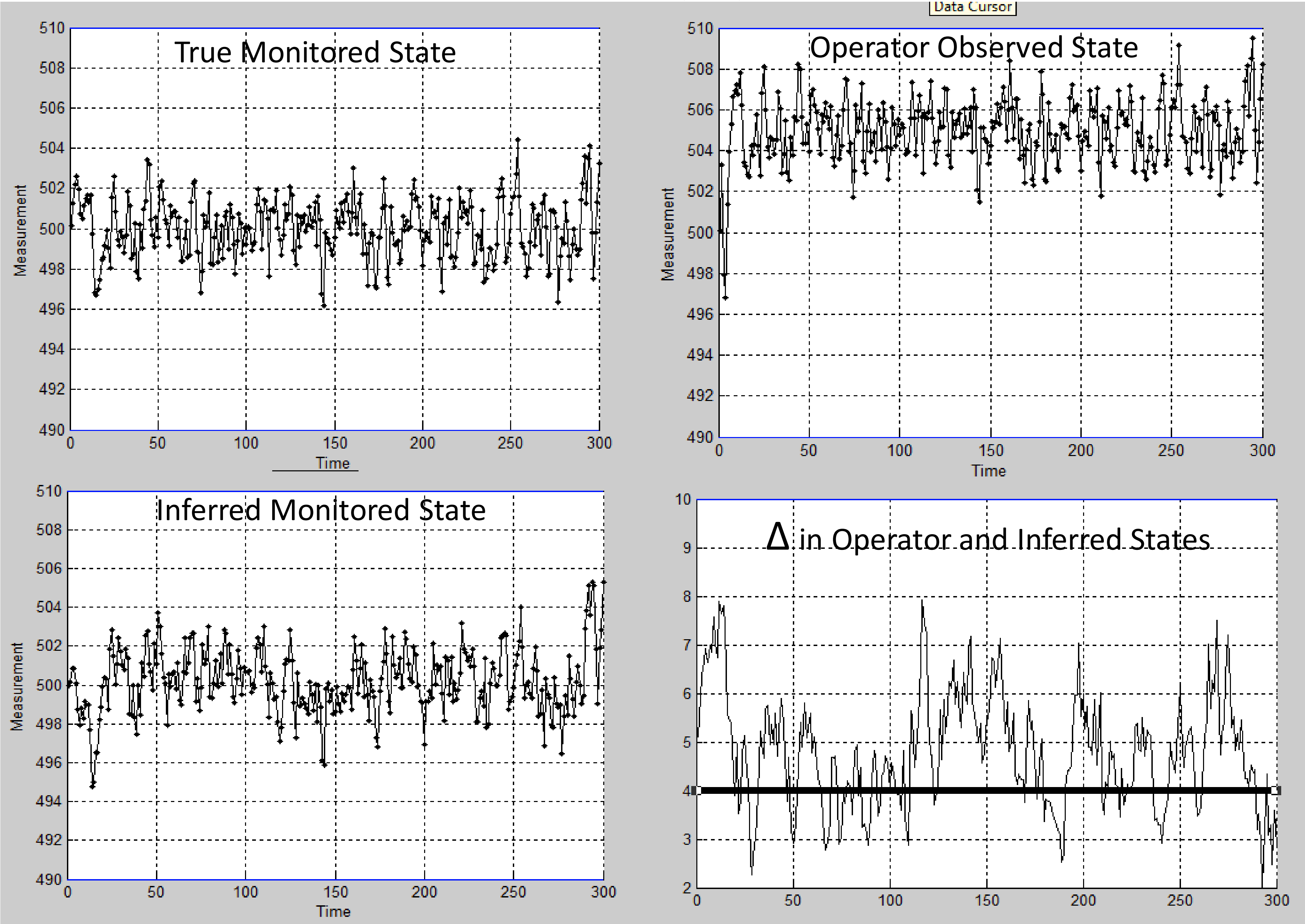
True Monitored State

Operator Observed State

Inferred Monitored State

Δ in Operator and Inferred States

Simulated Replay Attack



True Monitored State

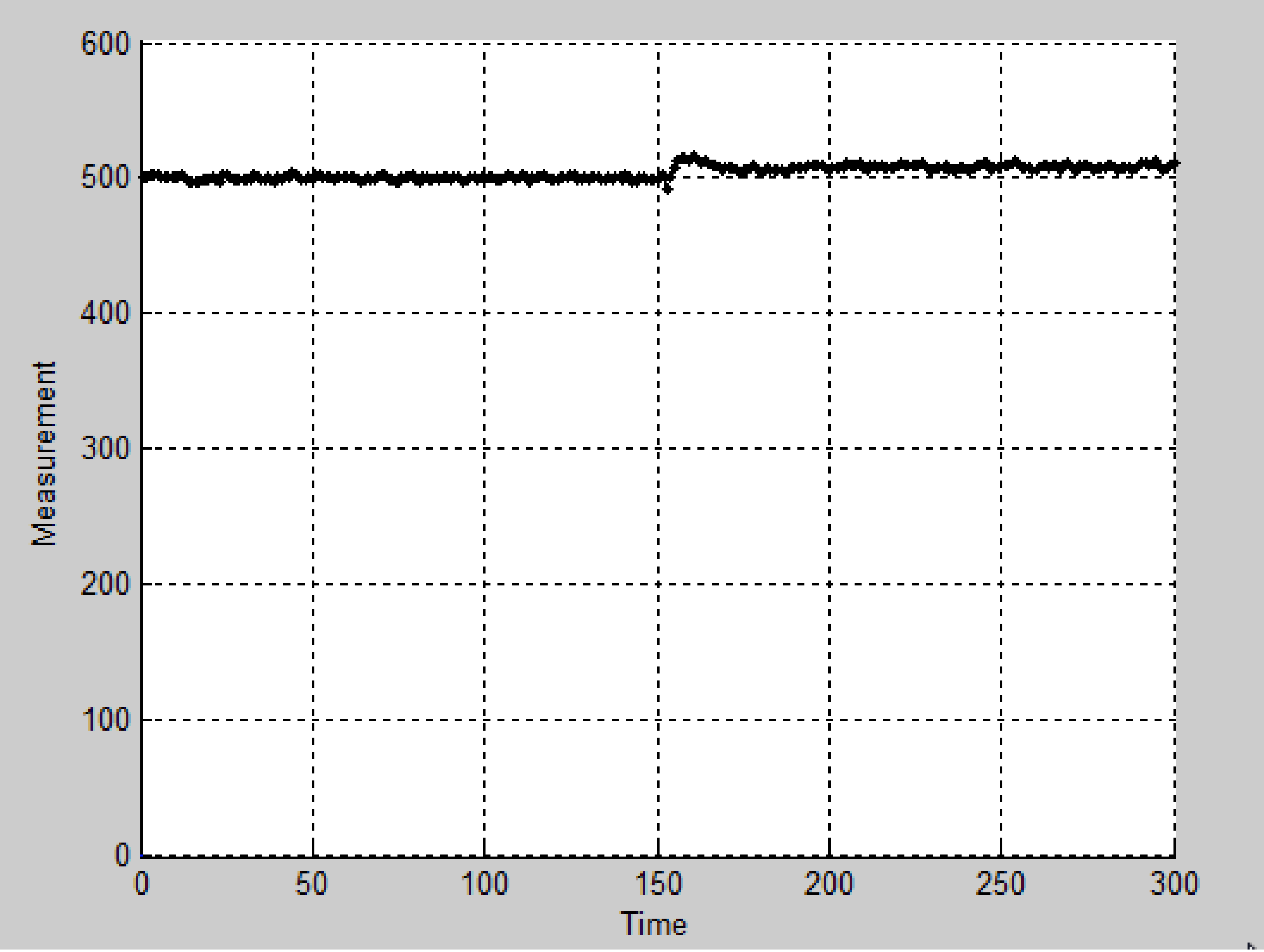
Operator Observed State

Inferred Monitored State

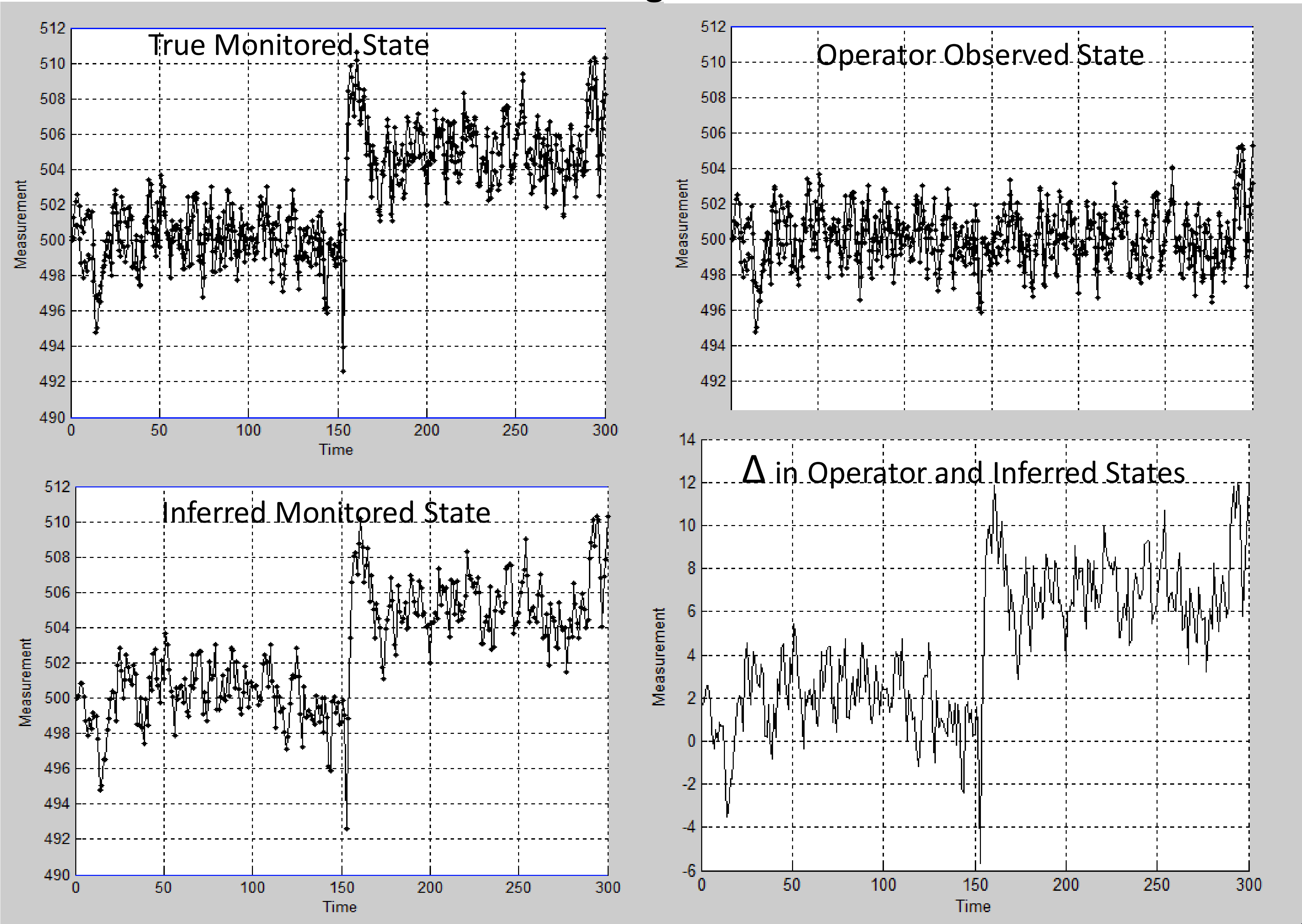
Δ in Operator and Inferred States

ATTACK TO ADJUST REGULATOR OBJECTIVES AND MASK THE PHYSICAL CHANGE THROUGH REPLAY ATTACK ON OPERATOR DISPLAYS

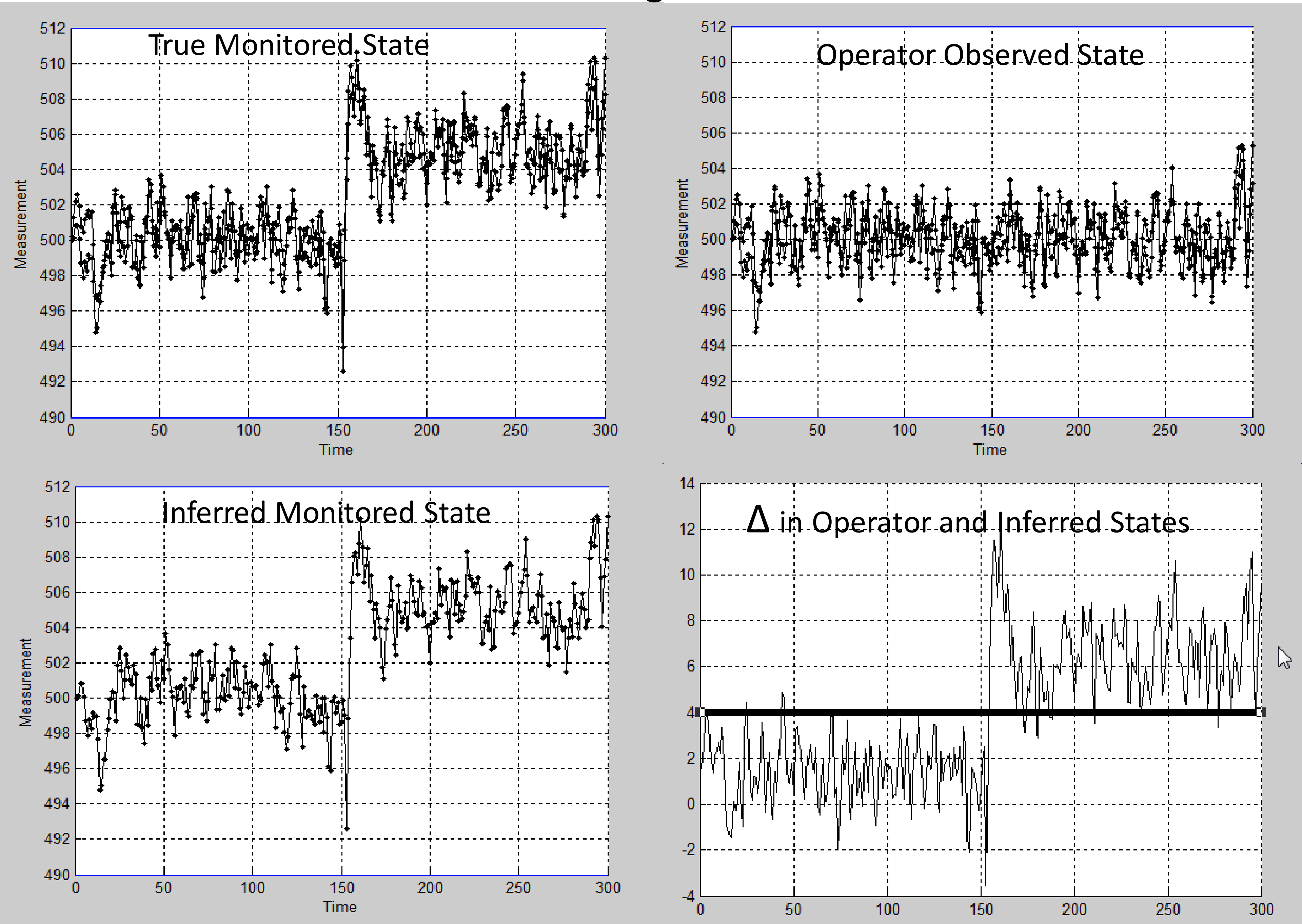
Simulated System Output Based Upon Controller Attack



Simulated Regulator Attack



Simulated Regulator Attack



Metrics

- As a practical matter, cyber attack detection/response for mission critical physical systems will need to be tuned to have virtually no model-predicable false alarms for initiating significant responses, such as shut down (for emphasis referred to as “zero” model-based false alarms), while also promising “zero” missed detections.

- Equivalently, sensor accuracy and corresponding detection algorithms must permit use of attack detection thresholds that are greatly distanced from both normal system operation and system operation regions that result in unacceptable consequences

- In order to determine detection thresholds and the corresponding false alarm and missed detection rates, operational data collections would need to be used to build upon model-based analysis, serving to account for shortfalls in system models.

- Detection algorithms and criteria that cause delays in initiating responses must account for how long a system can operate in a region of the state space before an important response is too late

Sliding Window Detection

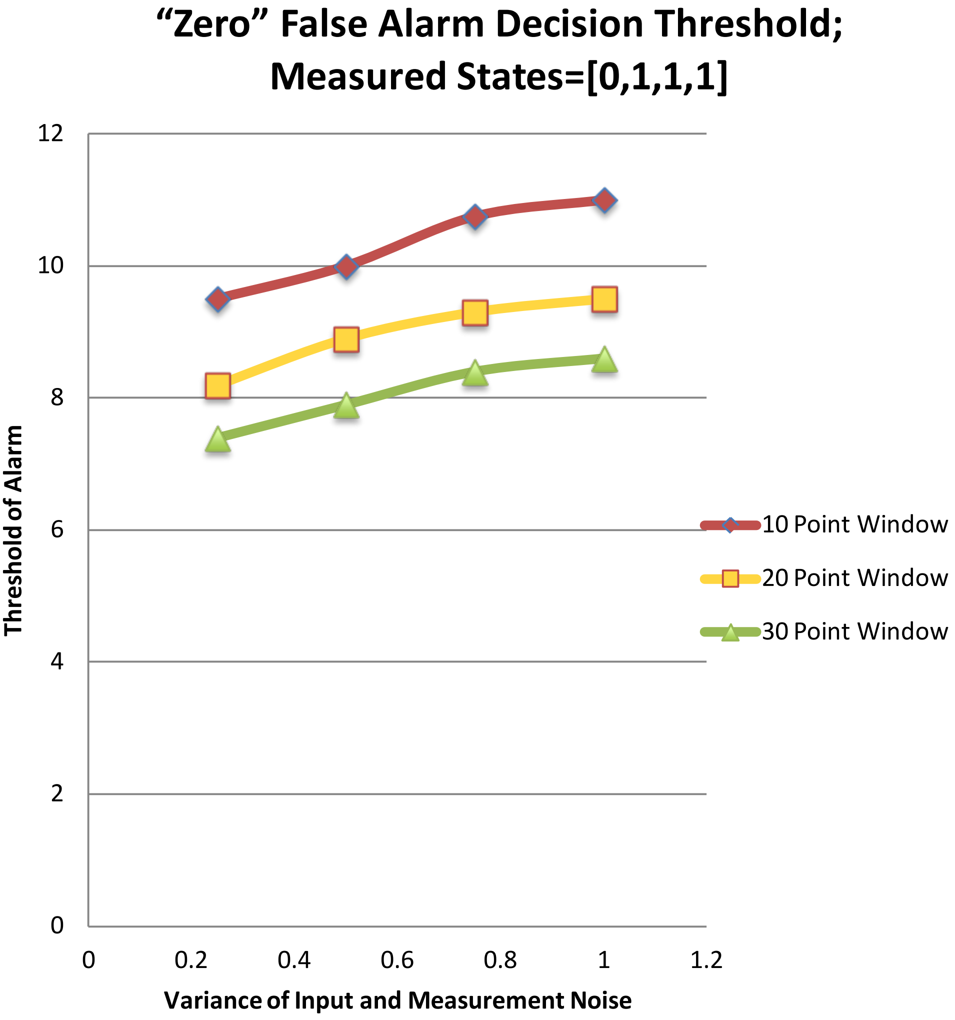
- For our example, a sliding window detection algorithm is used for integrating over the time series of the "N" most recent individual point detections, each based on a threshold test
  - A cyber attack is declared upon detecting m threshold violations over N detection opportunities
  - Increasing m and N serve to reduce over-reaction to individual estimates resulting in threshold violations, thereby reducing false alarm rate at the expense of potentially increasing the missed detection rate and delaying detections
- More specifically, given a time series of individual point detections, determined by comparing a time series of the most recent state estimates, x_1, x_2, x_3….x_N to an alarm threshold, ***th***
- If x_i> ***th***, increment g by 1, where:

$$g = \sum_{i=1}^{N} (x_i > \boldsymbol{th})$$

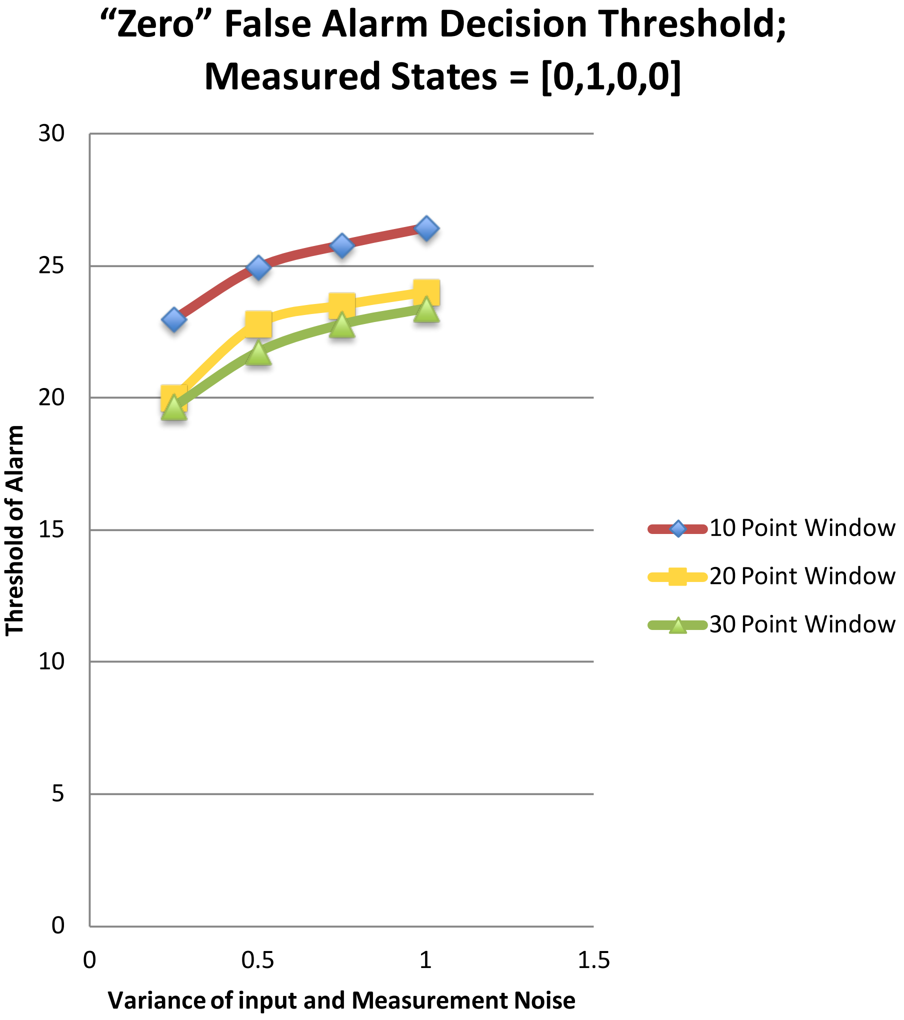
- For the example, within a time series consisting of N state estimates each compared to threshold criterion ***th***, if g > N/2 a cyber attack is declared.

“Zero” False Alarm Thresholds

150,000 point simulation



“Zero” False Alarm Decision Threshold; Measured States=[0,1,1,1]

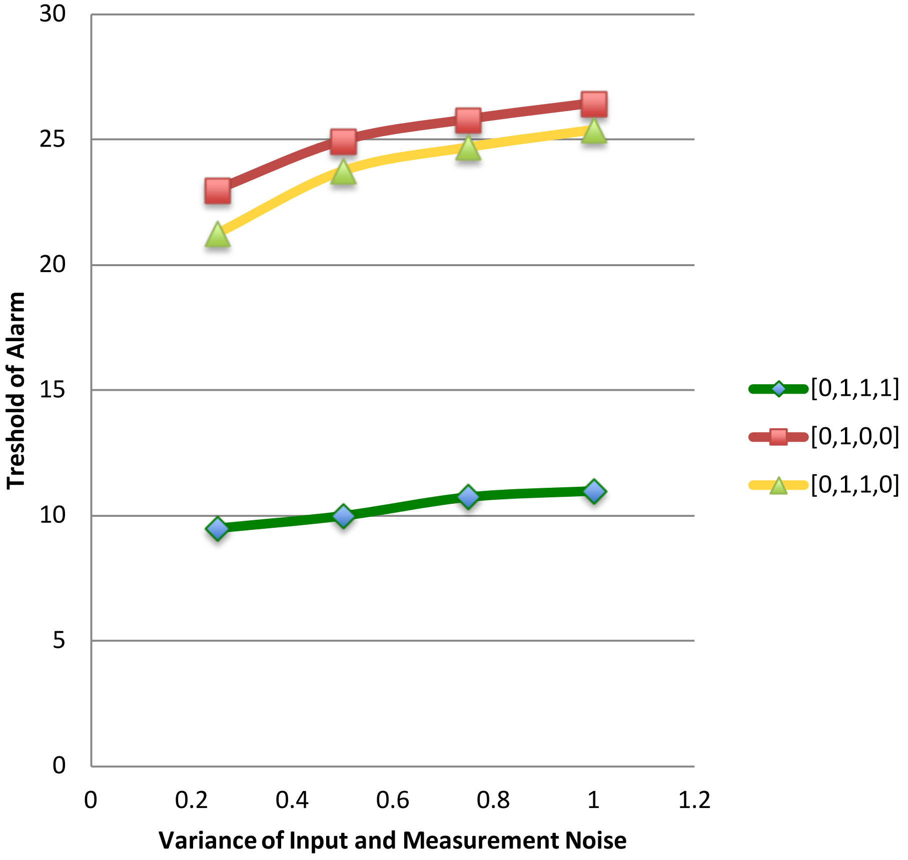


“Zero” False Alarm Decision Threshold; Measured States = [0,1,0,0]

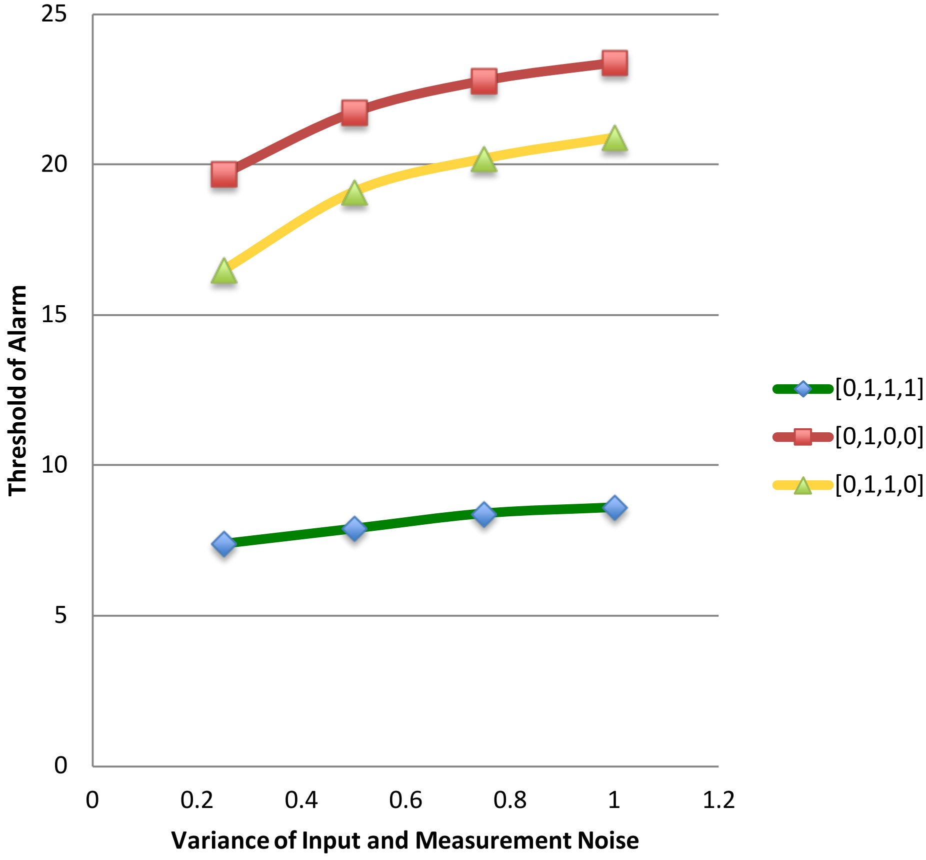
“Zero” False Alarm Thresholds

150,000 point simulation

“Zero” False Alarm Threshold; 10 Point Window / Minimum 10 Second Delay



“Zero” False Alarm Threshold; 30 Point Window/Minimum 30 Second Delay



Design Sensitivity Analysis

- Decision Thresholds vs sensor accuracy – ~20-30% change in threshold value over sensor accuracies (variances) ranging from 0.25 – 1
- Decision Thresholds vs selection of states used for inferring critical state(s) values – ~200-300% change in threshold value over state measurement range of [0,1,1,1] to [0,1,0,0]
- Decision Thresholds vs delays in detection (length of sliding window)-10-20% change in threshold value over a 10 – 30 second sliding window detector
- Design range of threshold values comparing the worst case (lowest thresholds) and best case designs (highest thresholds) for achieving "zero" model-based false alarm/missed detection rates – ~400% change from worst accuracy, least states measured, longest sliding window detector to best accuracy, most states measured, shortest sliding window detector

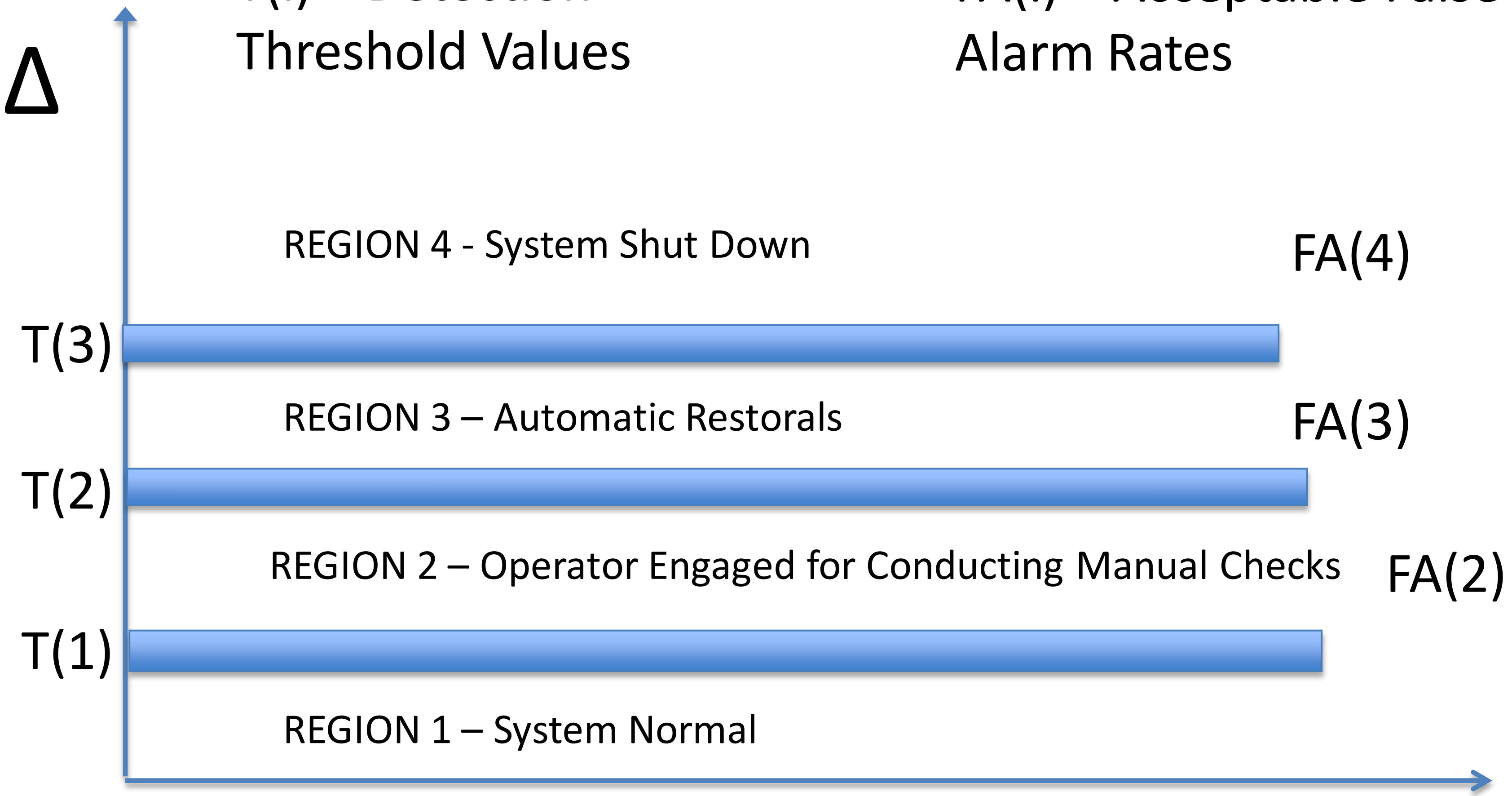
Real World Example: Gas Turbine

- RPM – 3600
- Measurement Error – 1-2 rpm ✔
- Data Interval - 40msec ✔
- Trip Threshold – ~10% rpm deviation ✔
- First estimate of augmenting sensor-based Trip Threshold - ~1% rpm deviation ✔
- Suitable spacing between attack detection thresholds and operating in regions with significant adverse consequences, permitting “zero” model-based false alarms/missed detections ✔
- Multiple triplex sensors – A/D converters and processor interfaces on a single board ✖

Relating Detection Thresholds, System Responses, and Acceptable False Alarm Rates

T(i) – Detection Threshold Values

FA(i) – Acceptable False Alarm Rates



Δ

REGION 4 - System Shut Down FA(4)

T(3)

REGION 3 – Automatic Restorals FA(3)

T(2)

REGION 2 – Operator Engaged for Conducting Manual Checks FA(2)

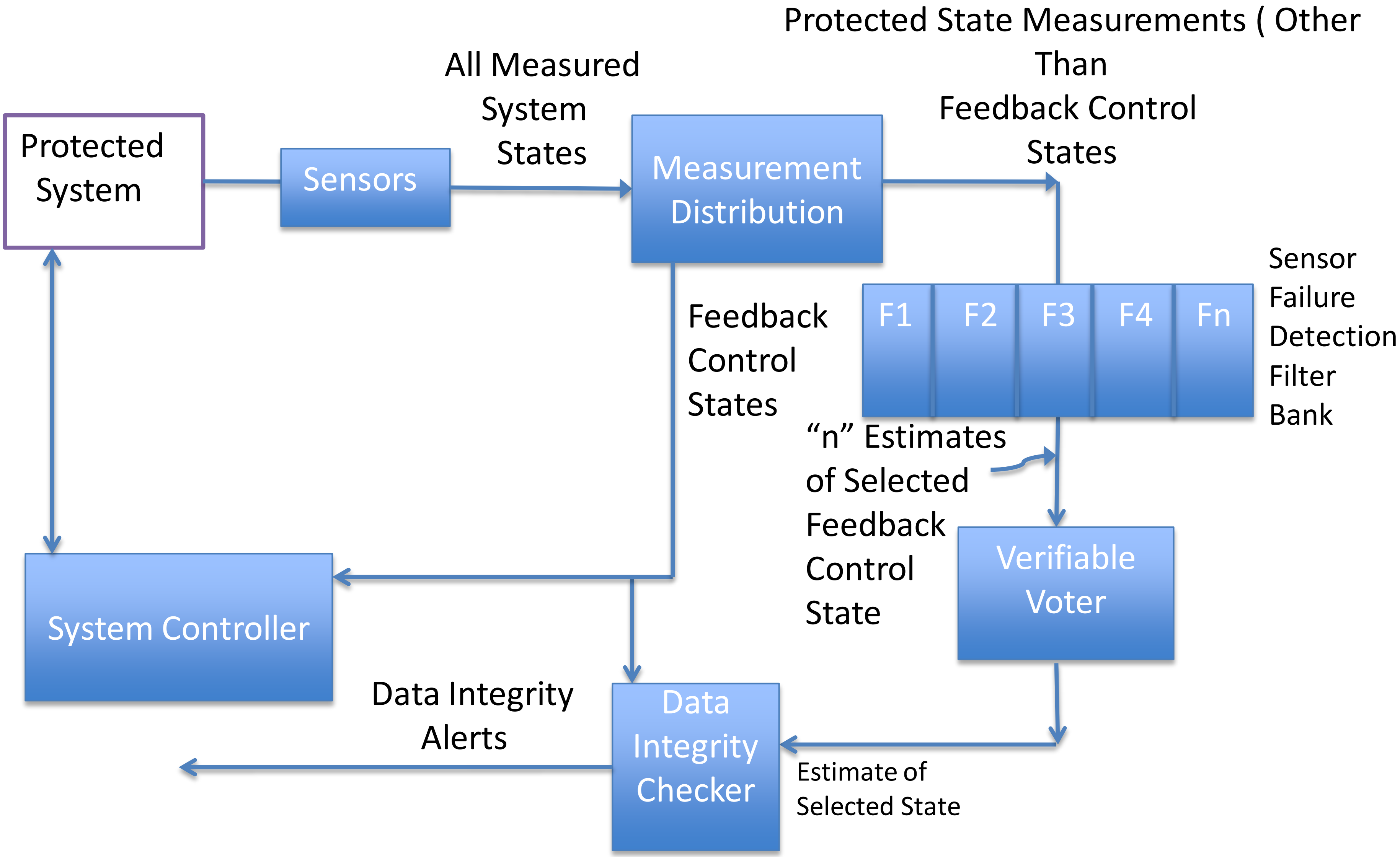
T(1)

REGION 1 – System Normal

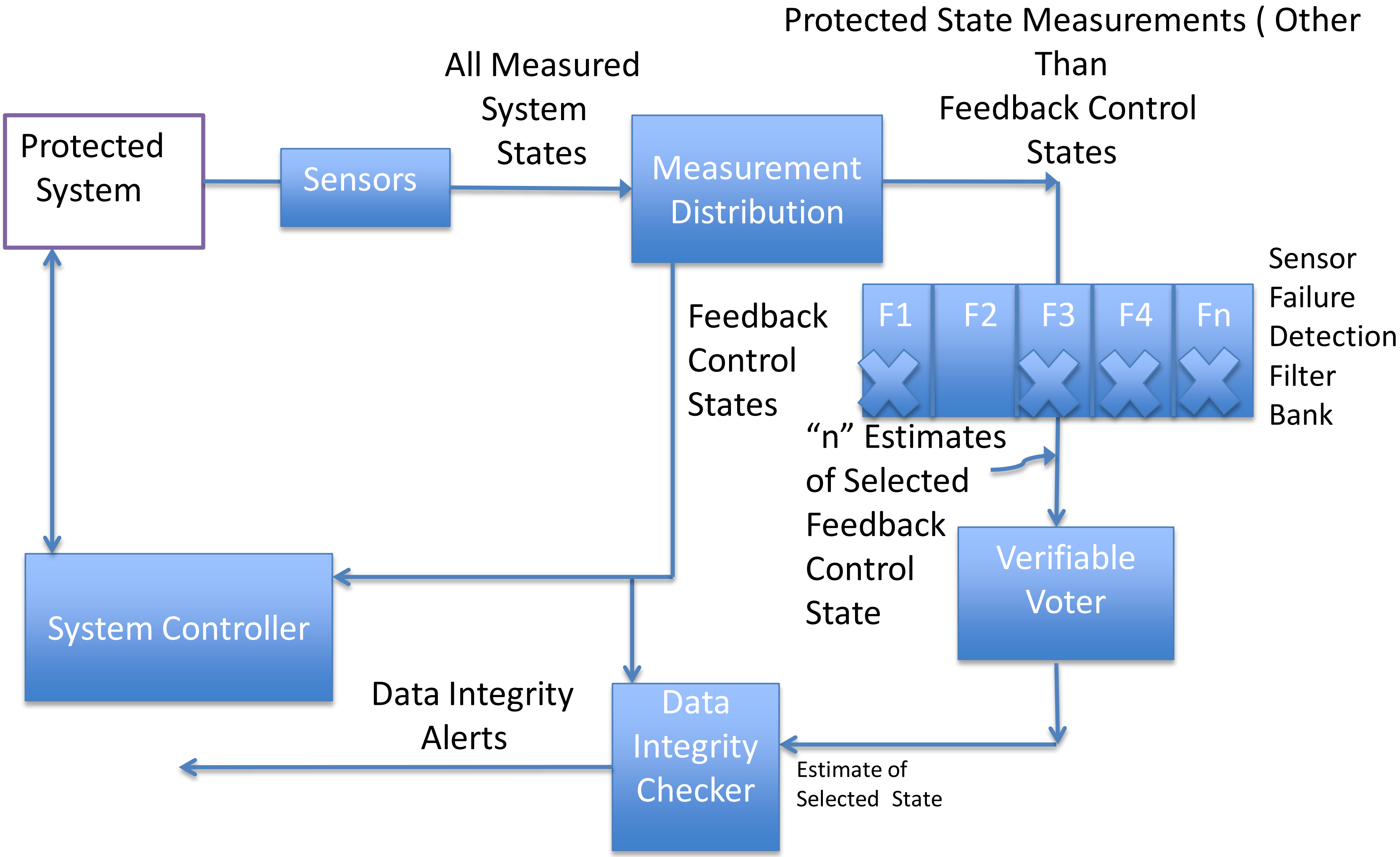
ATTACK ON CRITICAL SENSORS' OUTPUTS

Design Pattern Based Upon Cyber Security Extension of:
T. Kobayashi, D. L. Simon, Application of a Bank of Kalman Filters for Aircraft Engine Fault Diagnostics, Turbo Expo 2003, American Society of Mechanical Engineers and the International Gas Turbine Institute, June, 2003

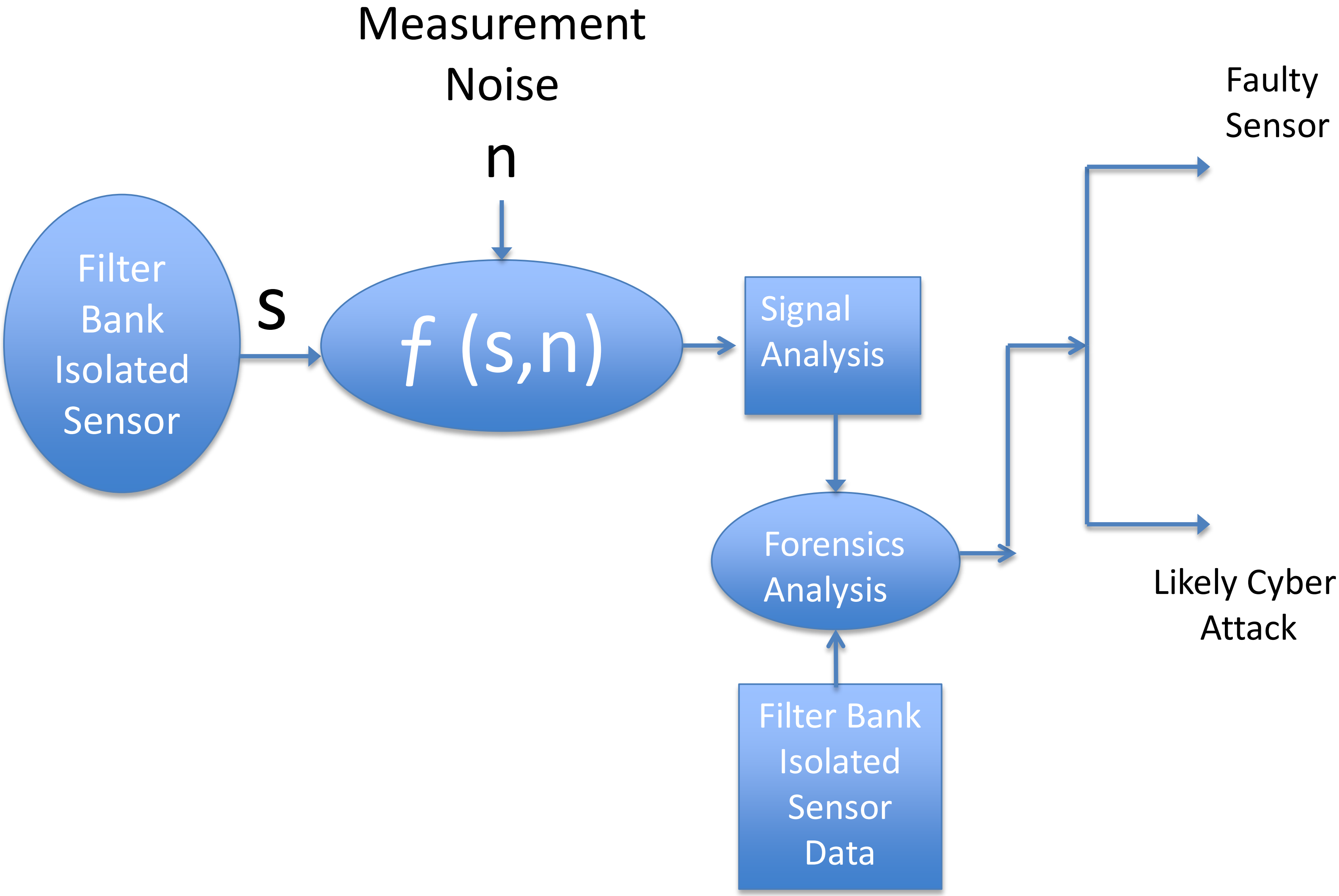
Simplified Block Diagram for Sensor Attack Detection System


Protected System
Sensors
All Measured System States
Measurement Distribution
Protected State Measurements (Other Than Feedback Control States
Sensor Failure Detection Filter Bank
F1
F2
F3
F4
Fn
Feedback Control States
"n" Estimates of Selected Feedback Control State
Verifiable Voter
System Controller
Data Integrity Alerts
Data Integrity Checker
Estimate of Selected State

Simplified Block Diagram for Sensor Attack Detection System


Protected System
Sensors
All Measured System States
Measurement Distribution
Protected State Measurements (Other Than Feedback Control States
F1
F2
F3
F4
Fn
Sensor Failure Detection Filter Bank
Feedback Control States
"n" Estimates of Selected Feedback Control State
Verifiable Voter
System Controller
Data Integrity Alerts
Data Integrity Checker
Estimate of Selected State

Rapid Post-Attack Sensor Noise Analysis to Confirm Faulty Sensor Assessment



Conclusions

- Data consistency checking design patterns can potentially make an important contribution to cyber security of physical systems

- Past work in fault-tolerant and automatic control systems provides a starting point regarding solutions and knowledge to draw upon, although specific solution designs will need to be implemented in a manner that is sensitive to the issues surrounding cyber attacks

- Development of actual solutions will require system activities in:
    - System dynamics modeling
    - State estimation
    - Security-focused analysis regarding attack scenarios, protection needs, more trusted and less trusted components, and sensors and measurement characterization
    - Distributed security solution designs that serve to complicate, and hopefully deter, attacks
    - In-field data collections regarding selection of detection thresholds and responses to achieve acceptably low false alarm/missed detection rates